

## ステンシル計算 FPGA アクセラレータのための時間 ・空間的並列度オートチューニングに関する研究

著者	遠藤 司
雑誌名	東北大学電通談話会記録
巻	88
号	1
ページ	244-245
発行年	2019-07
URL	<a href="http://hdl.handle.net/10097/00126634">http://hdl.handle.net/10097/00126634</a>

修士学位論文要約（平成31年 3 月）

# ステンシル計算 FPGA アクセラレータのための 時間・空間的並列度オートチューニングに関する研究

遠藤 司

指導教員：張山 昌論

## Automatic Tuning of Temporal and Spatial Parallelism for Stencil Computation FPGA Accelerators

Tsukasa ENDO

Supervisor: Masanori HARIYAMA

Recently, Field Programmable Gate Arrays (FPGAs) have attracted attention as high speed and low power consumption computing platforms for large-scale computations such as high-performance computing and big data processing. However, because the design freedom is very high, it is difficult to design an optimal architecture. Moreover, appropriate performance tuning is required according to a FPGA board. Therefore, in this research, we propose an automatic tuning method for temporal and spatial parallelism of FPGA accelerators. We implement the proposed method to find the optimized architecture for stencil computation. According to the results, the design time has been reduced to 3% ~ 11% of the conventional approach.

### 1. はじめに

近年、高性能計算・ビッグデータ処理といった大規模計算をより高速かつ低消費電力で処理できる計算プラットフォームとして、Field Programmable Gate Array(FPGA)が注目されている。FPGA は再構成可能な集積回路であり、演算対象に合わせて専用ハードウェアを作成できる。しかしながら、設計の自由度が非常に高いため、最適アーキテクチャ設計が難しい。また、使用する FPGA ボードに合わせて適切な性能チューニングが必要である。現在、C 言語ベースの設計環境(OpenCL for FPGA[1])を用いた高位合成によって、設計の簡単化が図られているが、最適設計は依然として難しく、多くの設計時間が必要である。そこで本研究では、OpenCL 設計手法を用いて FPGA アクセラレータの並列度を自動的にチューニングし、FPGA ボードに合わせて最適アーキテクチャが得られる設計手法を提案する。提案手法をステンシル計算アクセラレータ設計に適用することで、設計自動化を行い、設計時間を従来手法の 3~11%に削減できることを示す。

### 2. FPGA アクセラレータの性能予測手法

OpenCL コードのコンパイルには 2 段階の stage (1st と 2nd)が存在し、特に 2nd stage には数時間に及ぶ時間がかかる。そこで、提案手法ではコンパイルの 1st ステージで得られるレポートからコードの構造を解析し、FPGA 実行にかかるサイクル数を予測することで、短時間での並列度チューニングを行う。図 1 にレ

ポート解析から得られる OpenCL コードの構造例を示す。各ブロックは for ループ、レベルは階層、II(インシエーション・インターバル)はそのブロックが何サイクルごとに実行されるか、Iterations は実行回数、BI(Block Iterations)はIIと Iterations の積であり、各ブロックの総実行回数を示す。BI を同レベルにおけるブロックの依存関係に従って計算することで、コードの総サイクル数を求める。図 1 において、ブロック 2、3 に依存関係がないと仮定すると、同レベルのブロックは並列実行可能なため、レベル 1 の実行サイクル数はブロック 2 の 8N である。これが N 回実行されるため、ブロック全体の実行回数は 8N×N である。提案手法では、FPGA ボードによるリソース制約、外部メモリバンド幅制約条件下でコードの総サイクル数が最小になる並列度を自動的に探索する。

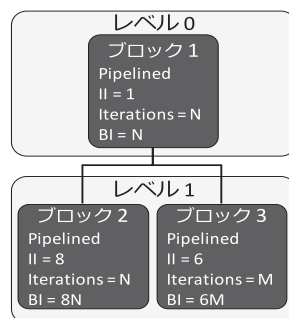


図 1 OpenCL コードの構造例

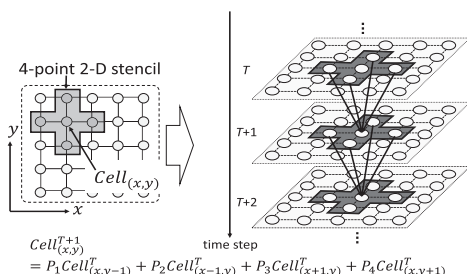


図 2 2D-4point ステンシルの計算

### 3. ステンシル計算アクセラレータのオートチューニング

ステンシル計算は、流体力学、電磁気シミュレーションなど、多くの分野で用いられている反復計算手法である。図 2 に 2D-4point ステンシルの計算を示す。新しいタイムステップにおけるセル(座標)の計算は、前のタイムステップにおけるセルの計算結果を使用して行われる。計算方法と計算量は適用するステンシルの形によって変化する。図 3 に先行研究[2]で提案されたステンシル計算アーキテクチャを示す。1 タイムステップを処理する複数のパイプライン演算モジュール

(PCM) で構成される。また、1 回のステンシル計算が 1 個の PE で実行され、複数の PE がステンシルを並列計算するために使用される。また、周波数と nPE が一定であれば、nPCM が大きいほうが高性能であることが保証されている。提案手法では、この前提条件と性能予測手法を用いて、ステンシルアプリケーションに最適な時間的並列度 PCM 数(nPCM)と空間的並列度 PE 数(nPE)を探索する。図 4 にオートチューニングの流れを示す。1st stage では制約内の各 nPE に対して、最小のクロックサイクル数となる nPCM を探索する。これらは最適解の候補となる。2nd stage では、候補すべてに 2nd stage コンパイルを行い、それぞれの処理時間を評価して最適解を決定する。

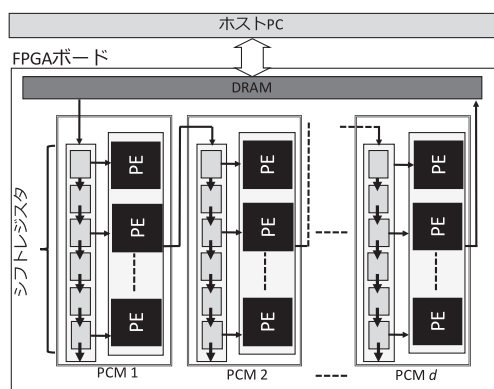


図 3 ステンシル計算アーキテクチャ

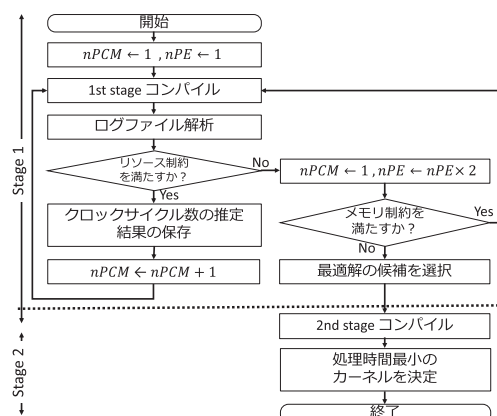


図 4 オートチューニングの流れ

### 4. 評価

表 1 に FPGA (DE5) を用いた場合の異なるステンシルアプリケーションに対する最適並列度と設計時間を示す。従来手法は制約内のすべてのパラメータに対して 2nd コンパイルを行った場合である。結果より、提案手法により各アプリケーションにおいて最適並列度を求め、設計時間を削減できた。

表 1: 異なるアプリケーションに対する最適並列度と設計時間(DE5)

アプリケーション	nPE	nPCM	従来手法[h]	提案手法[h]
Laplace	8	18	528	19.6
2-D 5-point	1	51	198.1	10
2-D 9-point	4	7	161.5	12.3

### 5. まとめ

本論文では、FPGA アクセラレータの性能予測手法とステンシル計算アクセラレータのオートチューニング手法を提案した。提案手法により、設計自動化を行い、設計時間を従来手法の 3 ~ 11% に削減できた。

### 文献

- 1) Czajkowski, T.S., Neto, D., Kinsner, M., Aydo- nat, U., Wong, J., Denisenko, D., Yiannacouras, P., Freeman J., Singh, D.P., Brown S.D., "OpenCL for FPGAs: Prototyping a compile" International Conference on Engineering of Reconfigurable Sys-tems and Algorithms (ERSA), pp.3-12, (2012)
- 2) Waidyasooriya, H.M., Takei, Y., Tatsumi, S., Hariyama, M., "OpenCL-Based FPGA-Platform for Stencil Computation and Its Optimization Method- ology" IEEE Transactions on Parallel and Dis- tributed Systems, Vol. 28, No. 5, pp.1390-1402, (2017)